



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### The General Linear Model: theory and practicalities in brain morphometric analyses

**Citation for published version:**

Pernet, CR 2018, The General Linear Model: theory and practicalities in brain morphometric analyses. in *Brain Morphometry: Methods and Clinical Applications*. Neuromethods, vol. 136, Springer, pp. 75-85.  
<[https://link.springer.com/protocol/10.1007/978-1-4939-7647-8\\_5](https://link.springer.com/protocol/10.1007/978-1-4939-7647-8_5)>

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Brain Morphometry: Methods and Clinical Applications

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Brain Morphometry: Methods and Clinical Applications (Springer - 2017)

---

*Editors: Gianfranco Spalletta, Tommaso Gili & Fabrizio Piras.*

## The General Linear Model: theory and practicalities in brain morphometric analyses (chapter 1.5)

---

*Cyril R. Pernet,*

*Centre for Clinical Brain Science, The University of Edinburgh*

### Abstract

The General Linear Model (GLM) is the statistical method of choice used in brain morphometric analyses because of its ability to incorporate a multitude of effects. This chapter starts by presenting the theory, focusing on modelling and then goes on discussing multiple comparisons issues specific to voxel based approaches. The end of the chapter discusses practicalities: variable selection and covariates of no interest. Researchers have often a multitude of demographic and behavioural measures they wish to use and methods to select such variables are presented. We end with a note of caution as the GLM can only reveal covariations between brain and behaviour, prediction and causation mandate specific designs and analyses.

*Key words:* General linear model, inference, variable selection, covariates.

This is the author pre-print, distributed under CC-BY.

The edited version is available as book chapter in the reference above.



## Introduction

At the core, the General Linear Model (GLM) comes from regression and correlational methods and it can be understood as a general multiple regression model. It derives from the theory of algebraic invariants developed in the 19<sup>th</sup> century along with the development of linear algebra. The theory searches for explicit description of polynomial functions<sup>1</sup> that do not change under transformations from a given linear group. Linear correlations for example are invariant, i.e. the linear correlation coefficient does not change after linearly transforming the data. For a thorough explanation of the GLM, see Christensen (2002). The GLM has been successfully used in the analysis of brain structures because of its flexibility to handle both categorical (e.g. groups of subjects), and continuous variables (e.g. test scores). It can be used to examine regions of interest (ROI) from which various morphometric markers can be extracted, but it has been most successful in whole brain analysis using a voxel based approach for grey matter (Voxel Based Morphometry or VBM, (Ashburner & Friston, 2000) and white matter (Tract Based Spatial Statistics or TBSS, Smith et al., 2006).

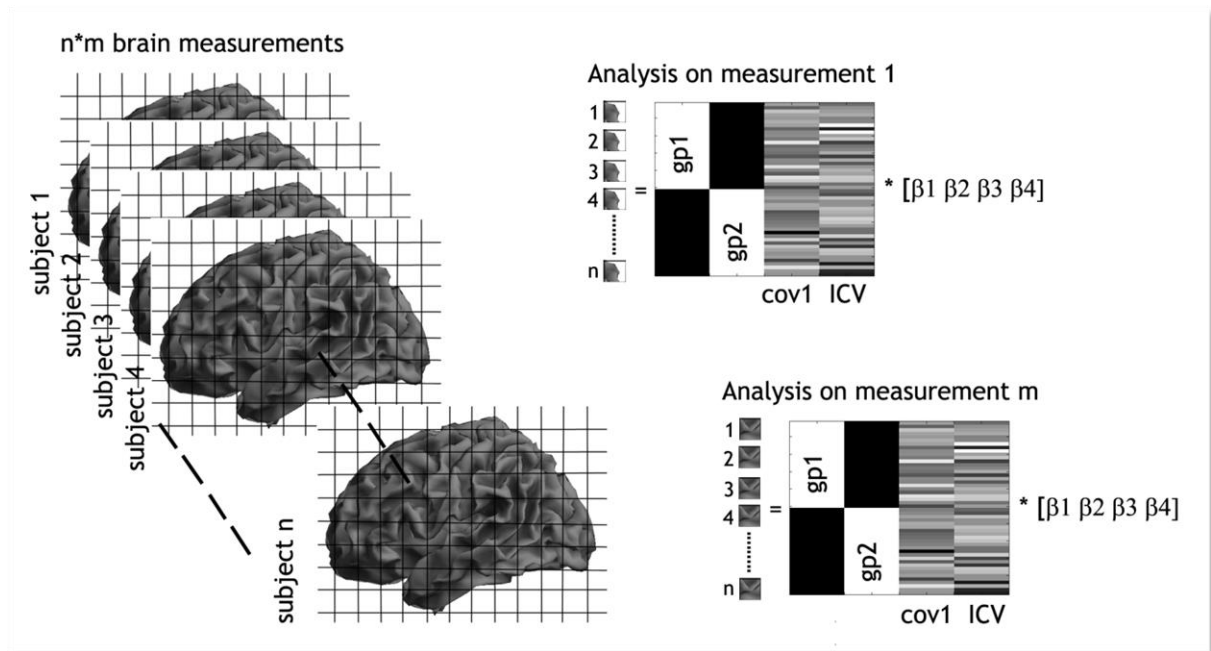
## Theory

The dependent variables (Y) can be described as a linear combination of independent variables (XB). Written as an equation, this gives  
$$Y = X\beta + \varepsilon \quad (\text{eq. 1}).$$

Each term is a matrix of data with Y of dimensions  $n \times m$ , X of dimensions  $n \times p$ ,  $\beta$  of dimension  $p \times m$  and  $\varepsilon$  of dimensions  $n \times m$  represents the residuals, i.e. the part of the data not explained by  $X\beta$ . Typically  $m=1$  when performing morphometric analyses (Figure 1), that is each region of interest (ROI) or each voxel of the brain is modelled separately (massive univariate approach). When using multiple ROI it can be useful to have a single model with  $m>1$ , taking advantage of correlational structure in the data to find a significant multivariate effect in the absence of significant univariate differences (e.g. patients can differ from controls by exhibiting lower volumes in regions A and B and higher volumes in regions C and D, but none of those regions show on their own a significant difference).

---

<sup>1</sup> A polynomial function is a function such as a quadratic, a cubic, a quartic, and so on, involving only non-negative integer powers of x.



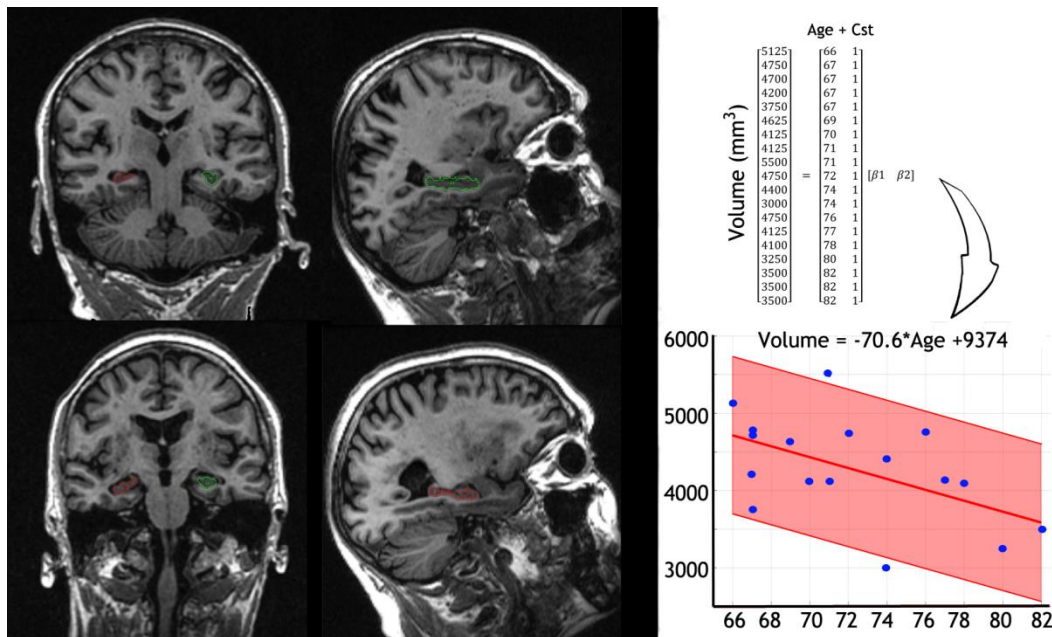
**Figure 1:** Illustration of the mass-univariate GLM approach. Data consist in  $n$  subjects and the brain of each subject is 'cut' into  $m$  measurements (voxels or ROI). Each measurement is then submitted to the same analysis, i.e. the same model  $X$  (here illustrated with a matrix of 4 columns, modelling membership of group 1 or 2, a covariate of interest and the Intra-cranial volume (ICV) as a controlling confounding variable). The parameters  $\beta$ , or linear combinations of them, are then tested for significance either for each of these measurements or for topological features associated to these measurements (i.e. clusters height or size).

## Linearity

The first aspect to recognize in equation 1 is linearity. Linearity refers to a relationship between variables that can be geometrically represented as a straight line (2D), a plane (3D) or a hyperplane (4D and above). A linear function satisfies the properties of additivity and scaling, which means that outputs ( $Y$ ) are simply an addition of inputs ( $X$ :  $x_1$ ,  $x_2$ ,  $x_3$ , etc.) each one of them multiplied by some constants ( $\beta$ :  $b_1$ ,  $b_2$ ,  $b_3$ , etc.).

Let us consider the application of such model to hippocampal volume measurements. I use here rounded data from Frisoni et al. (2008, figure 3) and look at the relation between healthy subject ages and normalized volumes of the right hippocampus<sup>2</sup>. The model is a simple linear regression such as  $y = x_1b_1 + b_2$ . The equation is linear since it states that  $y$ , the hippocampal volume, is equal to age ( $x_1$ ) scaled by  $b_1$  plus an average volume value ( $b_2$ ). Geometrically, the relationship is defined as a straight line (Figure 2).

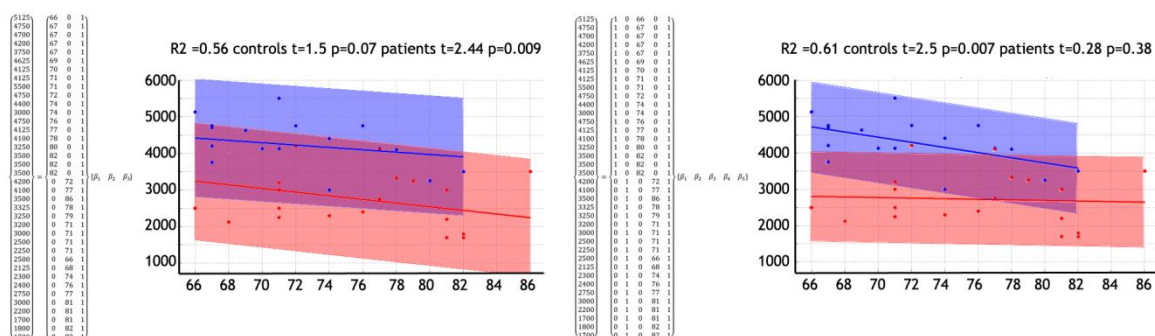
<sup>2</sup> In morphometric analyses, the total intracranial volume (or related measurement) is typically accounted for, either in the model or in the data. Here the hippocampal volume is normalized to the total brain volume – this transformation is mandatory as bigger heads give bigger volume and vice-versa and bias in a sample can lead to spurious results.



**Figure 2: Figure 3:** Simple regression on a morphological feature: here hippocampal volume. After delineating the hippocampus of 19 subjects, (image courtesy from the Centre for Clinical Brain Science, Edinburgh Imaging Library brain-DA0001), Frisoni et al. (2008) computed the normalized hippocampal volume and performed a regression analysis of the volume as a function of the subject's age (plus a constant term), showing a significant reduction of hippocampus volume as one gets older.

## Modelling data

The second aspect to recognize in equation 1 is the modelling. Consider again the regression model for the hippocampus volume, adding now the group of Alzheimer patients. We can either model the data with a single mean or having group specific means in addition of group specific age regressors. In the former case, we build a model in which controls and patients have overall the same volumes but we allow age related changes to differ between groups (i.e. the hypothesis is that aging does not affect the brain the same way between subjects who have or not the disease). In the latter case, we build a model in which controls and patients have different overall volume values (i.e. they come from different distributions, which is a reasonable assumption given what we know about Alzheimer and memory) while also exhibiting differential age related changes (Figure 3).



**Figure 4.** Two different models comparing the hippocampal volumes of healthy controls and Alzheimer patients. Including the group variable changes the slope of the age regressors, giving different results/conclusion about how grey matter plasticity changes with age.

It is important to recognize that any data analysis using the GLM is driven by hypotheses. Whilst this might seem trivial, differences between models can lead to large differences in results and authors and readers of scientific articles should be aware of this. Here both models are statistically significant (the 1<sup>st</sup> model has a  $R^2$  of 0.56 and the 2<sup>nd</sup> model a  $R^2$  of 0.61). The simple inclusion or exclusion of the group variable leads however to opposite conclusions. If we consider that both controls and patients come from the same distributions, patients show a significant correlation with age. If we consider group specific volumes, the correlation with age is only significant in control subjects (implying the disease 'destroyed' or obscured age related changes).

## Model Estimation

Once the model is set-up, a software solves the equation, i.e. it estimates the model parameters. If we consider a linear system, the solution is found simply by multiplying each side of eq. (1) by the inverse of  $X$  giving the solution in equation 2. Most often  $X$  is not square and the system is not fully determined (i.e. we don't have an exact set of equations that describes the data). The solution is therefore a set of estimates. Typically, the Moore-Penrose pseudoinverse (eq. 3) is used, but other solutions exist (e.g. QR decomposition, Bayesian estimation), which implies that slightly different results can sometimes be obtained because of the estimation method.

$$\hat{\beta} = X^{-1}Y \text{ (eq. 2)}$$

$$\hat{\beta} = X^+Y \text{ (eq. 3)}$$

## Hypothesis testing

Once the model has been set up and the parameters have been estimated, it is time for statistical testing. Testing always follow the same rule: effect / error (equation 4). One way to think about hypothesis testing is to think about how to combine parts of the model (i.e. the columns of  $X$ ) to explain the data ( $Y$ ). One can test if only one regressor explains the data, or if a set of regressors explains the data, or even if one regressor or set of regressors explains more of the data than another regressor or set of regressors. Contrasts are simple linear functions that combine the columns of  $X$  to test such hypotheses. Taking the model with 5 columns in Figure 3, we can test using equation 4 if there is significant effect of age in controls ( $C=[0 \ 0 \ 1 \ 0 \ 0]$ ) or in patients ( $C=[0 \ 0 \ 0 \ 1 \ 0]$ ) or even across the whole sample ( $C=[0 \ 0 \ 1 \ 1 \ 0]$ ). It is also possible to test if there is a difference in overall volumes between groups ( $C=[1 \ -1 \ 0 \ 0 \ 0]$ ) or a difference in the age effect ( $C=[0 \ 0 \ 1 \ -1 \ 0]$ ).

$$t = \frac{c\hat{\beta}}{\sqrt{\sigma^2 c^T (X^T X)^{-1} c}} \quad \text{eq (4)}$$

*C defined the contrast of interest,  $\hat{\beta}$  are the parameter estimates,  $\sigma^2$  is the variance obtained from the residuals,  $X$  is the design matrix.*

## Multiple comparisons correction

The type I error rate corresponds to probability to observe false positives, i.e. the probability to declare a ROI or voxel significant whilst  $H_0$  (the absence of effect) is true. Because  $m$  statistical tests are usually performed, we have to consider the family-wise error rate (FWER), i.e. the probability to make one or more type I errors, in order to ensure an overall error at the pre-specified alpha level. Many methods have been devised to control the type I FWER at the voxel or cluster level: Bonferroni correction, Random Field Theory (Worsley et al., 1996) and randomization tests (permutation or bootstrap – Hayasaka et al., 2004). Note that false discovery rate (FDR) procedures do not control the type I error in the classical sense. FDR procedures control for the expected number of false positives.

If testing multiple ROI, one of those techniques must be used: either FDR in the context of data exploration or FWER for hypothesis driven studies. While both Bonferroni and randomization tests will control the type 1 FWER to a similar level, Bonferroni correction is not optimal because it considers ROI as independent, whilst ROI data are often correlated. Randomization tests are likely to provide more power, accounting for correlations between ROI. With full brain analyses, i.e. testing many spatially contiguous voxels (i.e. VBM or TBSS), the choice of the multiple comparison technique has an even greater impact. It has been shown that statistical values are attracted toward regions of low residual variance – which should not happen if the data were stationary. Unfortunately full brain voxel based analyses suffers from non-stationarity and this means that in smooth regions, clusters tend to be large even in the absence of true signals, resulting in increased false positives whilst in rough regions, clusters tend to be small, resulting in reduced power. Because of this sensitivity to smoothness, it is not recommended to threshold maps based on cluster size (Ashburner & Friston, 2000) unless using specifically a non-stationary permutation cluster test (Hayasaka, Phan, Liberzon, Worsley, & Nichols, 2004). An interesting solution is to integrate size and height over all thresholds, a technique known as Threshold Free Cluster Enhancement (S. Smith & Nichols, 2009) which has been successfully applied to TBSS and is becoming more popular for VBM when, again, it is adjusted for non-stationarity (Salimi-Khorshidi et al., 2009, 2011). Finally, it is also worth considering the number of contrasts performed – the more covariates are tested, the more likely one of them will be significant by chance, and alpha level adjustment could also be performed at that stage (Ridgway et al., 2008 - rule 8).

## Practicalities

### Assumptions

The 1<sup>st</sup> job of the experimenter and data analyst is to define the independent variables. Sometimes people worry about the normality assumption. It is important to understand that this only applies to the data (and model residuals) and not to the regressors, which means that as long as subjects are independent and identically distributed (i.i.d), anything can be entered as regressors, for instance a quadratic effect of age. It is however worth also considering the relationships between variables as it is often not clear if brain structural changes are explained by the behaviour or if this is the behavioural data that are predicted by brain morphology (Huang et al., 2013).

**Normality.** When performing the analysis on a few ROI, it is advisable to check the data distributions as long tails will give spurious results. With VBM, grey matter values are bounded and it is thus recommended to have sufficient spatial smoothing, from 4 to 8mm (Salmond et al., 2002), to render



the distributions normal and thus avoid high false positive rate. In this context, unbalance between groups is the most problematic with extreme cases such as comparing a single subject to a group leading to high number of false positives (Scarpazza et al., 2013).

**Independence.** Lack of independence occurs when the error covariance is not diagonal. There is no good way to fix such problem because it does not relate to the data but to the sampling. The single best advice is thus to have a good sampling strategy for the population of interest.

## Choosing covariates

To investigate structural brain changes in health and disease, many variables of interest are obtained for each participant. For instance, researchers working on language disabilities measure basic sensory performances along with IQ, auditory and visual cognitive abilities. Similarly, research in aging will try to characterize many cognitive dimensions using multiple tests along with a more general test such as the MoCA (Sink et al., 2015). Of particular interest for GLM analyses are (i) the accuracy of these measures and (ii) the relationship between these measures.

All sensory and cognitive measures are affected by random measurement errors. This means that measurements fluctuate randomly around their true values because of inherent biological variability, of imprecise tool, or both. The total error in a variable with random measurement error averages out to zero, which implies that on average this is accurate. Many people assume that it implies that it does not affect regression analyses. Since GLM estimates minimize the square distance in  $X$ , errors in the regressors (i.e. the sensory or cognitive measurements) lead to poor data fit, biasing the regression coefficient towards the null, a phenomenon known as attenuation or regression dilution bias (Hutcheon et al., 2010). When we have multiple variables which measure the same sensory or cognitive dimension, best is to choose the one with the smallest error – i.e. the one with the smallest standardized standard deviation.

Having multiple variables that measure the same or overlapping dimensions can also be a problem for morphometric analyses. If the goal were to explain the data with a model, collinearity between regressors would not be an issue. Since the goal is usually to investigate how each variable relates to structural changes, this becomes a problem. Multicollinearity leads to a decrease in the unique amount of variance explained by a single regressor thereby posing difficulty in interpreting their respective contribution. Kraha et al., (2012) propose a set of tools that can be useful in this context (e.g. commonality analysis, dominance analysis), examining relationships between the regressors and the variable of interest. These tools are easy to apply for ROI analyses but are intractable for VBM or TBSS analyses. One useful tool is the variance inflation factor, which examines the impact of each regressor on the others, therefore making sure no regressors in the model are too much correlated with others or a combination of them. In many cases, the experimenter or data analyst will still have to select some variables. Without a tractable solution over all voxels, one option is to rely on a decomposition of behavioural data to build a new (smaller) set of regressors that reflect specific dimensions with minimal (factor analysis) to no correlation (principal component analysis) between them. Using such approaches has also the advantage to have predictors with minimized variance, i.e. likely reduced measurement errors.



## Common covariates of no interest

Depending on the research question at hand, several confounding variables can be included in the design matrix. A few other covariates, related to brain size, should also be included: age, sex, total intra-cranial volume (TIV). The exact relationships between global and local morphology to age and sex are not known, but these factors are known to affect grey and white matter tissues and thus should be included. Because morphometric analyses are interested in local changes, it is also recommended to control for the TIV and/or related measurements (Malone et al., 2015). For instance, we can include TIV to account for total head size or the total grey matter (TGM) or total white matter (TWM) to account for tissue-specific global effects. However, not all of these covariates are needed, and their inclusion/exclusion must be justified. As discussed in Henley et al. (2010), adjusting data using TIV eliminates differences due to sex, making that variable irrelevant. Similarly, adjusting for age and TGM, approximates TIV because age and TGM are correlated in normal subjects. Such choice is however invalid in neurodegenerative disease since TGM does not follow TIV. In all case, one global measure is likely to be included and there are different ways to perform this control (Peelle et al., 2012):

1. *Local Covariation*: the model assumes that observed changes are explained by the sum of a global and a local effect. The global effect (TIV, TGM or TWM) is however scaled differently at each voxel/ROI, which implies that one accepts the hypothesis that the global effect affects different regions differently. For instance, one brain region can be smaller as head size increases while another region can be bigger. This approach is computed by simply adding the global measure as covariate in the GLM. The results are interpreted as significant changes, in addition of, or despite changes in the global.
2. *Global Scaling*: the model assumes changes are explained by a local effect that scales beyond changes in the global measurement. Here, the global effect is accounted for identically for every voxel/ROI by simply dividing them by the global estimate. A significant effect of a regressor indicates changes that are stronger than changes in the global measurement. This approach is particularly useful when changes in TGM or TWM are known and/or compete with the local effect tested like in e.g. aging.
3. *Local Scaling*: the model assumes changes are explained by a local effect that scales beyond changes in the global measurement as for global scaling. This effect however scales differently at each voxel/ROI as in the local covariation approach. This is computed by dividing the data by the local mean values. Compared to the local covariation approach, this allows looking for effects that scales beyond global changes, rather than in addition of the global effect. Since the adjustment is local, it seems that part of the effects observed (compared to the local covariation) can be attributed to partial volumes rather than uniquely attributed to a global confound.

In general local covariation have shown to be more effective to control for the global effect of head size (see references in Malone et al., 2015) but scaling methods allow to answer slightly different questions and have therefore merits of their own. No matter what model is used, it is important to check effects with and without covariates of no interest (in particular TIV or related), to better understand observed changes (Tu et al., 2008). Good reporting practices are then to (i) say if an effect is there with/without the covariates, (ii) report standardized regression coefficient  $\text{std}(X)/\text{std}(y) \cdot \beta$  along with statistical values, allowing quantitative comparison across studies and, (iii) also report changes in terms of effect size (e.g. raw gray matter values) to provide a more direct biological interpretation.

## Discussion

As presented above, the GLM is a mathematical tool of great flexibility allowing one to combine categorical (e.g. patients vs. controls) and continuous variables (e.g. age, TIV), to test multiple hypotheses within the same model thus accounting for confounding effects. Results from such an approach must however be considered carefully because (i) of the assumptions of the normality, of independence, and of linearity (see e.g. Pernet, et al., 2009) for non-linear effects) (ii) multiple models are often possible and can give different results, i.e. the model reflects the experimenter's hypotheses and assumptions about the population(s), (iii) significant effects of regressors are not always good biomarkers or predictors. This last point is of utmost clinical importance. The GLM is useful to understand how brain structures linearly covaries with behavioural measures but it does not provide information of causation or diagnosis. Causation can only be established by design, as for instance in a randomized controlled interventional study (pre- post- treatment with placebo), showing that changes in behaviour after treatment corresponds to changes in brain structure in treated patients but not placebo. Similarly, to establish that a behavioural variable or a brain region is a good biomarker, cross-validation techniques must be used and a new independent subject sample must be tested to validate the prediction because significant GLM regressors are not necessarily good predictors (Lo et al., 2015).

## Acknowledgments

Thank you to Ged Ridgway for providing useful references related to variables of no interest and reviewing the manuscript, and to David Raffelt for pointing out difference between looking at voxel content (VBM/TBSS) versus morphometry per se (looking at shapes).

## References

- Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*, 11(6), 805–821. <http://doi.org/10.1006/nimg.2000.0582>
- Christensen, R. (2002). *Plane Answers to Complex Questions. The theory of Linear Models*. (3rd ed.). New-York: Springer.
- Frisoni, G. B., Ganzola, R., Canu, E., Rub, U., Pizzini, F. B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C., & Thompson, P. M. (2008). Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. *Brain : A Journal of Neurology*, 131(Pt 12), 3266–3276. <http://doi.org/10.1093/brain/awn280>
- Hayasaka, S., Phan, K. L., Liberzon, I., Worsley, K. J., & Nichols, T. E. (2004). Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage*, 22(2), 676–687. <http://doi.org/10.1016/j.neuroimage.2004.01.041>
- Henley, S. M. D., Ridgway, G. R., Scahill, R. I., Kloppel, S., Tabrizi, S. J., Fox, N. C., & Kassubek, J. (2010). Pitfalls in the use of voxel-based morphometry as a biomarker: examples from huntington disease. *AJNR. American Journal of Neuroradiology*, 31(4), 711–719. <http://doi.org/10.3174/ajnr.A1939>
- Huang, L., Rattner, A., Liu, H., & Nathans, J. (2013). How to draw the line in biomedical research. *eLife*, 2. <http://doi.org/10.7554/eLife.00638>
- Hutcheon, J. A., Chiolerio, A., & Hanley, J. A. (2010). Random measurement error and regression dilution bias. *BMJ (Clinical Research Ed.)*, 340, c2289.

- Kraha, A., Turner, H., Nimon, K., Zientek, L. R., & Henson, R. K. (2012). Tools to Support Interpreting Multiple Regression in the Face of Multicollinearity. *Frontiers in Psychology*, 3. <http://doi.org/10.3389/fpsyg.2012.00044>
- Lo, A., Chernoff, H., Zheng, T., & Lo, S.-H. (2015). Why significant variables aren't automatically good predictors. *Proceedings of the National Academy of Sciences*, 112(45), 13892–13897. <http://doi.org/10.1073/pnas.1518285112>
- Malone, I. B., Leung, K. K., Clegg, S., Barnes, J., Whitwell, J. L., Ashburner, J., Fox, N., & Ridgway, G. R. (2015). Accurate automatic estimation of total intracranial volume: A nuisance variable with less nuisance. *NeuroImage*, 104, 366–372. <http://doi.org/10.1016/j.neuroimage.2014.09.034>
- Peelle, J. E., Cusack, R., & Henson, R. N. A. (2012). Adjusting for global effects in voxel-based morphometry: Gray matter decline in normal aging. *NeuroImage*, 60(2), 1503–1516. <http://doi.org/10.1016/j.neuroimage.2011.12.086>
- Pernet, C. R., Poline, J., Demonet, J., & Rousselet, G. A. (2009). Brain classification reveals the right cerebellum as the best biomarker of dyslexia. *BMC Neuroscience*, 10(1), 67. <http://doi.org/10.1186/1471-2202-10-67>
- Ridgway, G. R., Henley, S. M. D., Rohrer, J. D., Scahill, R. I., Warren, J. D., & Fox, N. C. (2008). Ten simple rules for reporting voxel-based morphometry studies. *NeuroImage*, 40(4), 1429–1435. <http://doi.org/10.1016/j.neuroimage.2008.01.003>
- Salimi-Khorshidi, G., Smith, S. M., & Nichols, T. E. (2009). Adjusting the neuroimaging statistical inferences for nonstationarity. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2009* (pp. 992–999). Springer. Retrieved from [http://link.springer.com/chapter/10.1007/978-3-642-04268-3\\_122](http://link.springer.com/chapter/10.1007/978-3-642-04268-3_122)
- Salimi-Khorshidi, G., Smith, S. M., & Nichols, T. E. (2011). Adjusting the effect of nonstationarity in cluster-based and TFCE inference. *NeuroImage*, 54(3), 2006–2019. <http://doi.org/10.1016/j.neuroimage.2010.09.088>
- Salmond, C. H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D. G., & Friston, K. J. (2002). Distributional Assumptions in Voxel-Based Morphometry. *NeuroImage*, 17(2), 1027–1030. <http://doi.org/10.1006/nimg.2002.1153>
- Scarpazza, C., Sartori, G., De Simone, M. S., & Mechelli, A. (2013). When the single matters more than the group: Very high false positive rates in single case Voxel Based Morphometry. *NeuroImage*, 70, 175–188. <http://doi.org/10.1016/j.neuroimage.2012.12.045>
- Sink, K. M., Craft, S., Smith, S. C., Maldjian, J. A., Bowden, D. W., Xu, J., Freedman, B., & Divers, J. (2015). Montreal Cognitive Assessment and Modified Mini Mental State Examination in African Americans. *Journal of Aging Research*, 2015, 872018. <http://doi.org/10.1155/2015/872018>
- Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., ... Behrens, T. E. J. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4), 1487–1505. <http://doi.org/10.1016/j.neuroimage.2006.02.024>
- Smith, S., & Nichols, T. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1), 83–98. <http://doi.org/10.1016/j.neuroimage.2008.03.061>
- Tu, Y.-K., Gunnell, D., & Gilthorpe, M. S. (2008). Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon--the reversal paradox. *Emerging Themes in Epidemiology*, 5, 2. <http://doi.org/10.1186/1742-7622-5-2>
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4(1), 58–73. [http://doi.org/10.1002/\(SICI\)1097-0193\(1996\)4:1<58::AID-HBM4>3.0.CO;2-O](http://doi.org/10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O)